

Nejjednodušší způsoby vyhledávání v Českém národním korpusu

Je třeba mít zajištěný plný přístup k Českému národnímu korpusu (dále jen ČNK)

I. Přístup k ČNK:

1. vyhledání webové stránky: www.korpus.cz
2. přihlášení se pomocí jména a hesla
3. přejít do webového rozhraní KonText (vlevo nahoře)

Objeví se strana, z které je možno v korpusu vyhledávat – viz rámeček **Hledat v korpusu**.

Korpus: v šedém rámečku vyberte z nabízených možností korpus **SYN2015**

Typ dotazu: volíme šipkou vpravo v bílém rámečku: teď nás zajímají **základní, lemma a slovní tvar** (viz dále).

II. Vyhledávání podle atributů **základní** (angl. *basic*), **lemma**, **slovní tvar** (angl. *word*)

1. Vyhledávání podle atributu **základní**

(a) Vybereme-li si v bílém rámečku šipkou vpravo atribut **základní** a zadáme-li do dlouhého bílého řádku **slovo v základním tvaru** (infinitiv slovesa, 1. pád sg. podstatného jména, 1. pád sg. maskulina přídavného jména...) a „klikneme“ na pokyn **Hledat** (v modrém rámečku dole), **vyhledá manažer KonText vždy všechny tvary daného slova**.

Příklady:

- do dlouhého bílého řádku napíšeme slovo *ležet*, KonText vyhledá všechny tvary tohoto slovesa;
- do dlouhého bílého řádku napíšeme slovo *kočka*, KonText vyhledá všechny tvary tohoto podstatného jména;
- do dlouhého bílého řádku napíšeme slovo *červený*, KonText vyhledá všechny tvary tohoto přídavného jména, tj. i tvary ženského a středního rodu.

(b) Vybereme-li si v bílém rámečku šipkou vpravo atribut **základní** a zadáme-li do dlouhého bílého řádku **slovo v jiném než základním tvaru**, např. *leží, ležíš, kočku, kočky, červená, červeného* a „klikneme“ na pokyn **Hledat** (v modrém rámečku dole), **vyhledá manažer KonText vždy jen uvedený slovní tvar** (tj. jen tvary *leží, ležíš, kočku, kočky, červená, červeného*).

2. Vyhledávání podle atributu **lemma**

Lemma je souhrn všech tvarů daného slova. Vybereme-li si v bílém rámečku šipkou vpravo atribut **lemma**, musíme do dlouhého bílého řádku zadat vždy jen základní tvar slova (slovesný infinitiv, 1. pád sg. substantiva, 1. pád sg. mužského rodu adjektiva).

Pozor! Pokud při volbě atributu **lemma** zadáme např. slovo *leží, kočku, kočky, červená, červeného*, **manažer KonText nevyhledá nic**.

3. Vyhledávání podle atributu **slovní tvar**

Vybereme-li si v bílém rámečku šipkou vpravo atribut **slovní tvar**, manažer KonText vždy vyhledá jen ten tvar slova, který zadáme, tj. „chová“ se tak, jako když zadáme jiný než základní tvar při volbě atributu „základní“. Ovšem pokud při volbě atributu slovní tvar do dlouhého bílého řádku zadáme např. slovo *ležet*, KonText vyhledá pouze infinitiv *ležet* v kladném tvaru, při zadání *kočka* KonText vyhledá pouze tvar *kočka* apod.

Vyhledávání pomocí atributů **slovní tvar** a **lemma** má proti vyhledávání pomocí atributu **základní** výhody, o nichž si řekneme příště.

III. V semináři jsme zjišťovali:

A. Poměr výskytu variantních tvarů slovesa pít v 1. osobě singuláru: *(ne)piji* : *(ne)piju*

Postup:

Vyhledáváme pomocí atributu **slovní tvar**.

1. a) Zvolíme atribut **slovní tvar** a do příkazového (tj. dlouhého bílého) řádku vepíšeme *piji* a stiskneme („odklikneme“) příkaz Hledat. KonText vyhledá 45 dokladů s tvarem *piji*.

b) Do příkazového řádku vepíšeme záporný tvar *nepiji* a stiskneme („odklikneme“) příkaz Hledat. KonText vyhledá 20 dokladů s tvarem *nepiji*.

Varianta s koncovkou „i“ v 1. osobě sg. má tedy 65 dokladů.

2. a) Zvolíme atribut **slovní tvar** a do příkazového (tj. dlouhého bílého) řádku vepíšeme *piju* a stiskneme („odklikneme“) příkaz Hledat. KonText vyhledá 276 dokladů s tvarem *piju*.

b) Do příkazového řádku vepíšeme záporný tvar *nepiju* a stiskneme („odklikneme“) příkaz Hledat. KonText vyhledá 143 dokladů s tvarem *nepiju*.

Varianta s koncovkou „u“ v 1. osobě sg. má tedy 419 dokladů.

Poměr variant <i>(ne)piji</i> : <i>(ne)piju</i> je tedy 65 : 419, tj. přibližně 1 : 6 (spíše 1 : 6,5).
--

B. Poměr výskytu pravopisných variant *filozofie* : *filosofie*.

– Při volbě atributu **lemma** a zadání *filosofie* do příkazového řádku KonText nevyhledá nic.

– Při volbě atributu **základní** a zadání *filozofie* do příkazového řádku KonText vyhledá pouze tvar *filosofie*, tj. nikoliv tvary *filosofii*, *filosofií*, *filosofiím*, *filosofiích*.

Proto zvolíme následující postup:

1. Zvolíme atribut **lemma** a do příkazového řádku zadáme *filozofie* (variantu se „z“) a stiskneme příkaz **Hledat**. KonText vyhledá 3905 dokladů, mezi nimiž budou jak podoby se „z“ (*filozofie*, *filozofií*...), tak podoby se „s“ (*filosofie*, *filosofií*...).

2. Přistoupíme k využití příkazu Frekvence – jednoho z nejdůležitějších při práci s korpusem:

- dotkneme se kurzorem příkazu Frekvence, objeví se nabídka „lemmata“, „slovní tvary“ a další: stiskneme **slovní tvary**;

- Kontext vytvoří seznam slovních tvarů („wordů“) seřazených podle frekvence. Abychom měli lehčí práci, stiskneme slovo „word“ nad frekvenčním seznamem: tvary *filozofie*, *filosofie*, *Filozofie*, *filosofií* apod. se uspořádají abecedně.

- Zjistíme, že méně je tvarů se „s“. Spočítáme (sečteme) jejich frekvence a dostaneme číslo 1062. To je výskyt všech variant se „s“ v celém korpusu SYN2015.

- Počet variant se „s“ i „z“ je 3905. Od tohoto počtu odečteme 1062, dostaneme počet variant se „z“ = 2843.

Poměr variant <i>filozofie</i> : <i>filosofie</i> je tedy $2843 : 1062 = \text{cca } 3 : 1$ (přesněji asi $2,7 : 1$).
--

Zvolte některý z postupů u A a B i při zpracovávání úkolu na příští seminář.